Synopsis for Registration as Research Student
For
Doctoral Degree
On
Satellite Image Analysis for Land Use Land Cover Application

By
Patil Parmeshwar Shyamsundar
Research Scholar

Supervisors

Dr. R.S.Holambe
Dr. L.M.Waghmare

Department of Electronics and Telecommunication Engineering
Shri Guru Gobind Singhji Institute of Engineering and Technology,
Nanded-431606 (MS.) India.
July 2019

1. Abstract

A stark increase in the amount of satellite imagery available in recent years has made the interpretation of this data a challenging problem at scale. Deriving useful insights from such images requires a rich understanding of the information present in them. thus posing huge challenges in automatic land cover (LC) and land use (LU) application. Deep learning reignited the pursuit of artificial intelligence towards a general-purpose machine to be able to perform any human-related tasks in an automated fashion. Several difficulties, however, are encountered when trying to apply the standard pixel-wise Convolution Neural Network for LC and LU application using satellite images, including geometric distortions, boundary uncertainties and huge computational redundancy.

These technical challenges for Land Use- Land Cover applications were solved by designing an automated framework for change detection and extracting semantic maps of roads, buildings, highways, vegetation etc. to track urban growth of cities in satellite images for land use land cover application purpose. Devising it as a supervised machine learning problem, a deep neural network is designed, implemented and experimentally evaluated. Publicly available datasets and frameworks are used for this purpose. The resulting pipeline includes image pre-processing algorithms that allows it to cope with input images of varying quality, resolution and channels. Therefore, deep learning techniques achieved by far the highest accuracy in the LC and LU applications, up to the existing deep learning methods, and traditional pixel-based and object-based approaches. This research made a significant contribution in Land Use and Land Cover information through deep learning-based innovations and has great potential utility in a wide range of geospatial applications.

Keywords: Deep Learning, land cover application, land use application, very fine spatial resolution, remotely sensed imagery, hierarchical representations.

2. Introduction

In order to make informed decisions pertaining to the environment, we are equipped with senses that allow us to observe it. This enables us to make effective changes around us as appropriate or desirable. We notice a need for understanding complex phenomena around the world such as urban growth, climate change, biodiversity studies and socioeconomic trends. This process very generally is referred to as earth observation, and has applications in disaster response, resource management and precision farming among others. Earth observation data is gathered by a range of techniques and can be roughly categorized as remote sensing and aerial photos.

Most of the ground features were captured manually through on-site survey and aerial photo interpretation, which are extremely labour-intensive and time-consuming. Arguably, this large archive of aerial imagery is highly under-utilised and could be mined for much more information efficiently and effectively through modern geospatial artificial intelligence (AI) and machine learning.

The most highly urbanised countries around the world developed over thousands of years of human habitation and are still rapidly changing with fast-paced and poorly planned urban growth, posing grand challenges across the country from environmental degradation for unsustainable economy. Such demanding problems, caused by rapid urban development, require responsive plans and decisions from environmental planners, policy makers, and local government authorities. Accurate and up-to-date land cover and land use (LULC) information is required therefore, urgently needed to keep pace with the ever-changing urban environments and to support relevant policies and decision-making.

Land use (LU), in comparison with Land Cover (LC), is less explored due to the complexity in spatial composition and configuration, where the land use patterns are formed by high-level semantics or functions. For example, land cover objects can be recognised as buildings, grassland, woodland etc. based on the low-level feature descriptors like spectra, texture, shape etc. whereas land use features are characterised as functional types with high-level semantics, such as residential, commercial, and industrial areas.

Recent advances in Artificial Intelligence and machine learning, especially the emerging field of deep learning, have changed the way we process, analyse and manipulate geospatial sensor data. LC and LU image application have not been explored until now, and the LU image information has not yet been solved due to the huge intra-class heterogeneity and inter-class similarity of land use features. Deep learning methods are naturally a good fit to capture such kinds of feature representations with high-level semantics.

Over the past few years, deep learning and deep convolutional neural networks (CNNs) have gained significant attention in the image analysis community. Different ground features occur across different scales, and their mapping objectives are strongly dependent upon the applications of interest. Until now, it is still an open question how to appropriately adopt or develop CNN-based methods to solve the complex LC and LU application tasks using remotely sensed and aerial imagery.

3. Literature Survey

The last decades have seen a constellation of sensors borne by diverse Earth observation platforms, including satellites, aerial systems and unmanned aerial vehicles (Othman et al. 2017). For Land Cover, traditional approaches can be divided broadly into pixel-based and object-based methods depending on the basic processing units, either per-pixel or per-object (Salehi et al. 2012).

The major challenges in applying these object-based approaches are the selection of segmentation scales to obtain objects that correspond to specific land cover types, in which over-segmentation and under-segmentation commonly exist within the same image (Ming et al. 2015).

Deep learning offers a different outlook on feature learning and representations, where robust, abstract and invariant features are learnt end-to-end, hierarchically, from raw data e.g. image pixels to semantic labels, which is a key advantage in comparison with previous state-of-the-art methods (Nogueira et al. 2017).

Many deep learning-based methods have been proposed, including deep belief networks (DBNs) (Chen et al. 2015), deep Boltzmann machines (DBMs) (Qin et al. 2017), stacked autoencoder (SDE) (Yao et al. 2016), and deep convolutional neural networks (CNNs) (Maggiori et al. 2017).

Amongst them, the CNN model represents the most well-established method, with impressive performance and great success in the field of computer vision and pattern recognition (Schmidhuber 2015), such as for visual recognition (Krizhevsky et al. 2012, Farabet et al. 2013), image retrieval (Yang et al. 2015) and scene annotation (Othman et al. 2016).

Deep learning is taking off in the field of remote sensing (Zhu et al. 2017). These publications show huge potential and practical utility in several remote sensing tasks, such as object detection (Zhao et al. 2017), semantic segmentation (Fu et al. 2017), road extraction (Wei, et al. 2017), and land use scene classification (Liu, et al. 2018).

Deep CNNs are a variant of multilayer neural networks that are specifically designed to process large-scale images or sensory data in the form of multiple arrays by considering local and global stationary properties (LeCun et al. 2015).

In the remote sensing domain, the equivalent problem is to solve the remotely sensed scene classification task, in which an image patch is assigned to a specific category, such as ‘airport’, ‘residential’, ‘commercial’. These sorts of land use scene classification tasks are closely related to object detection (Zhang et al. 2016) and localisation (Long et al. 2017).

where the translational invariance is the key advantage of the CNN to *detect* the object with higher order features, such as land use or functional sites. However, this characteristic becomes a major weakness in LC and LU application for pixel-level differentiation, from which blurred boundaries are produced between ground surface objects. Here, we review the satellite image analysis of both LC and LU application using CNNs to elaborate these challenges in detail and identify the research gaps.

Summary of the deep learning for solving remote sensing tasks with the number of papers and example sources Since 2014.

Remote Sensing tasks	**No. of Papers**	**Source (example)**
Target detection	22	Chen et al. 2014, Zhang et al. 2016, Long et al. 2017, Pei et al. 2018
Road extraction	6	Wang et al. 2015, Panboonyuen et al. 2017, Wei et al. 2017
Image processing	4	(Masi et al. 2016, Wei, Yuan, et al. 2017, Wang et al. 2018)
Semantic segmentation	49	Paisitkriangkrai et al. 2016, Maggiori et al. 2017, Wu et al. 2018
Land cover image classification	8	Kussul et al. 2017, Pan and Zhao 2017, Zhang et al. 2018
Land use scene classification	72	Hu et al. 2015, Othman et al. 2016, Nogueira et al. 2017, Liu et al. 2018
Change detection	5	Zhang et al. 2016, Gong et al. 2017, Su et al. 2017

From the table there are seven types of major applications. the remote sensing domain: target detection, road extraction, image processing, semantic segmentation, land cover image classification, land use scene classification and change detection. Among the seven applications, the land use scene classification (72 papers), semantic segmentation (49 papers) and target detection (22 papers) constitute the majority cases, whereas others are less researched so far Instead, the purpose is to provide a concise overview of deep learning methods for LC and LU application using remotely sensed imagery. We focus on deep convolutional neural networks (CNN), as they are the most typical and well-established deep learning method that has been adopted in the remote sensing domain.

4. Research Objectives

The main objective of this research is to produce the most accurate land cover and land use maps that are most suitable for meeting the potential customer requirements for Government Survey, such as to identify and understand land use changes in a fast-changing urban environment. To reach this aim, the following specific objectives are,

- Undertake a brief study on deep neural network techniques for very fine spatial resolution remotely sensed imagery.
- Develop a deep learning method that takes in data to produce land use-land cover maps on the remotely sensed images.
- Develop a deep learning method to solve the complex land use-land cover maps using remotely sensed imagery.
- Compare different neural networks on state-of-art dataset for land use-land cover application.
- Evaluate the trained network on a different dataset to understand how well it generalizes.

5. Methodology to be adopted

A primer on the current approach is provided here. The first presents a naive CNN implementation, followed by a description of the Fully Convolutional Network architecture, a part of which was used to extract features from a pretrained model.

5.1 Pre-processing

First task in the work is to collect authentic dataset i.e. with suitable resolution and from authentic agency. In literature we come to know that, at least 30m spatial resolution satellite images are suitable for proper land use land cover information. We also come to know that multi temporal and multisensory imagery is required. Understanding an image and classifying its content into semantic groups translates into formulating a per-pixel classifier, where we predict a class for each pixel in the image and extract a semantic map of the entire image. The same idea can be extrapolated into serving for a multi-class classification, where one would consider semantic groups such as buildings, meadows and rivers etc. Optical imagery is contaminated with cloud and shadows that's reason pre-processing is required. In this part missing values in the image has to restore by using various pre-processing methods and image segmentation and subsequent restoration of missing data in time series of satellite imagery is done.

5.2 Naive Approach

The initial model developed was a simple low-level layer architecture, with the primary aim of understanding how to implement a NN and get it running. For this, Dataset split into subset of training, validation and testing dataset, and later trained the low-level architecture on the entire strong training set. It is common to increase the number of feature maps produced in each layer as we go higher and higher.

Therefore, the network architecture consists of different convolutional layers with 32, 64, 128 feature maps respectively at each layer, followed by a dropout of some percentage. ReLU activations can be used in all the layers. Finally feature maps here is passed through a softmax function to obtain our probability heatmaps. All filters sizes followed the simplest possible structure 1x1, 3x3, 5x5, 7x7 sized kernels at each layer, stride and padding followed by kernel on each layer of architecture.

5.3 Transfer Learning

Many recent developments in the machine learning and computer vision area have been driven primarily using common benchmarks – models trained and tested on standard datasets of high variance, that generally lend themselves well to powerful features. The use of transfer learning allows one to use an existing model that has learned generalizable weights trained on a large dataset such as ImageNet, and fine tune the network to suit the particular use case. Convolutional networks learn features hierarchically. The generic descriptors that are obtained from a ConvNet hence provide a powerful starting point for fine-tuning existing models to a more specific task.

The advantage of transfer learning is discarding the fully connected final layers is that the spatial information - crucial in our context, is retained. In fully connected layers, each neuron is pairwise mapped to every single neuron of its preceding layer, regardless of spatial position. In stark contrast, a convolutional layer connects only to the neurons in its effective receptive field in a deep network.

5.4 Evaluation

Given any input, a binary classifier predicts either of two outcomes positive, or negative. For the pixel classification problem here, road pixels are considered positive, and background pixels, negative. The classifier output can be listed as:

- True Positives (TP): Foreground pixels are classified correctly.
- True Negatives (TN): Background pixels are classified correctly.
- False Positives (FP): Background pixels are mistakenly classified as roads.
- False Negatives (FN): Foreground pixels are mistakenly classified as the background.

These numbers are generally represented in a confusion matrix, as shown below, and the standard metrics used to evaluate a classifier are derived as different ratios from it. The models are then evaluated with these metrics on the initially isolated test data. This gives us a proxy measure of how well the model can generalize.

Two metrics relevant here are Precision and Recall, which are defined as:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

The area under a ROC curve provides a means of measuring the classifier's ability to discriminate between classes in the dataset. By this definition, maximizing the area (ROC) leads to a better classification accuracy. This metric is provided for comparison, and care must be taken to note that the datasets we deal with are highly imbalanced. Finally, the best performing models on each dataset are assessed with the F1-score and Intersection Over Union (IOU). These are defined as follows:

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

$$IoU = \frac{TP}{TP + FP + FN}$$

The F1 score is the harmonic mean of Precision and Recall, while IOU is interpreted as the name suggests: the intersection of the actual and predicted values, divided by the union of this set for a specific class.

6. Expected Outcome:

It is expected outcomes from the research work that:

1. Undertake a deep learning method for land use- land cover application using very fine spatial resolution remotely sensed images (VFSR).

2. Developed deep learning methods or architectures as a learning process towards human interpretation on VFSR remotely sensed imagery.

3. Solving complex LC and LU image information with a set of innovations through technological integrations and model iterations.

4. Understanding the segmentation, classification, change detection hierarchies and their intrinsic relationships across different mapping objectives, from simple land covers to complex land use feature representations.

5. Model the uncertainty in deep learning for very fine resolution remote sensed images for geographical survey application.

7. Refrences

1. Othman, E., Bazi, Y., Melgani, F., Alhichri, H., Alajlan, N., and Zuair, M., 2017. Domain Adaptation Network for Cross-Scene Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55 (8), 4441–4456.

2. Demarchi, L., Canters, F., Cariou, C., Licciardi, G., and Chan, J. C. W., 2014. Assessing the performance of two unsupervised dimensionality reduction techniques on hyperspectral APEX data for high resolution urban land-cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87 (July), 166–179.

3. Chen, Y., Lin, Z., Zhao, X., Wang, G., and Gu, Y., 2014. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7 (6), 2094–2107

4. Salehi, B., Zhang, Y., Zhong, M., and Dey, V., 2012. A review of the effectiveness of spatial information used in urban land cover classification of VHR imagery. *International Journal of Geoinformatics*, 8 (2), 35–51.

5. Ming, D., Li, J., Wang, J., and Zhang, M., 2015. Scale parameter selection by spatial statistics for GeOBIA: Using mean-shift based multi-scale segmentation as an example. *ISPRS Journal of Photogrammetry and Remote Sensing* [online], 106, 28–41.

6. Nogueira, K., Penatti, O. A. B., and dos Santos, J. A., 2017. Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognition* [online], 61, 539–556.

7. Chen, Y., Zhao, X., and Jia, X., 2015. Spectral-Spatial Classification of Hyperspectral Data Based on Deep Belief Network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8 (6), 2381–2392.

8. Qin, F., Guo, J., and Sun, W., 2017. Object-oriented ensemble classification for polarimetric SAR Imagery using restricted Boltzmann machines. *Remote Sensing Letters*, 8 (3), 204–213.

9. Yao, X., Han, J., Cheng, G., Qian, X., and Guo, L., 2016. Semantic Annotation of High-Resolution Satellite Images via Weakly Supervised Learning. *IEEE Transactions on Geoscience and Remote Sensing*, 54 (6), 3660–3671.

10. Maggiori, E., Tarabalka, Y., Charpiat, G., and Alliez, P., 2017. Convolutional Neural Networks for large-scale remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55 (2), 645–657.
11. Schmidhuber, J., 2015. Deep Learning in Neural Networks: An Overview. *Neural Networks*, 61, 85–117.

12. Krizhevsky, A., Sutskever, I., and Hinton, G. E., 2012. ImageNet classification with deep Convolutional Neural Networks. *In*: *NIPS2012: Neural Information Processing Systems*. Lake Tahoe, Nevada, 1–9.

13. Yang, X., Qian, X., and Mei, T., 2015. Learning salient visual word for scalable mobile image retrieval. *Pattern Recognition* [online], 48 (10), 3093–3101.

14. Othman, E., Bazi, Y., Alajlan, N., Alhichri, H., and Melgani, F., 2016. Using convolutional features and a sparse autoencoder for land-use scene classification. *International Journal of Remote Sensing* [online], 37 (10), 2149–2167.

15. Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., and Fraundorfer, F., 2017. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, 5 (4), 8–36.

16. Zhao, W., Du, S., and Emery, W. J., 2017. Object-Based Convolutional Neural Network for High-Resolution Imagery Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10 (7), 3386–3396.

17. Fu, G., Liu, C., Zhou, R., Sun, T., and Zhang, Q., 2017. Classification for High Resolution Remote Sensing Imagery Using a Fully Convolutional Network. *Remote Sensing* [online], 9 (5), 498.

18. Wei, Y., Wang, Z., and Xu, M., 2017. Road Structure Refined CNN for Road Extraction in Aerial Image. *IEEE Geoscience and Remote Sensing Letters*, 14 (5), 709–713.

19. Liu, Q., Hang, R., Song, H., and Li, Z., 2018. Learning Multiscale Deep Features for High-Resolution Satellite Image Scene Classification. *IEEE Transactions on Geoscience and Remote Sensing* [online], 56 (1), 117–126.

20. LeCun, Y., Bengio, Y., and Hinton, G., 2015. Deep learning. *Nature* [online], 521 (7553), 436–444.

21. Long, Y., Gong, Y., Xiao, Z., and Liu, Q., 2017. Accurate object localization in remote sensing images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing* [online], 55 (5), 2486–2498.

22. Zhang, P., Gong, M., Su, L., Liu, J., and Li, Z., 2016. Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116, 24–41.